Can Robots be "Self-Determining" ?

Sawa Kato

Summary

The topic of this thesis is "Self-determination in robots." First, I will distinguish the concept of self-determination from the "right" to self-determination. With this distinction in mind, two questions arise; one is whether robots can be self-determining, the other is whether robots have rights, including the right to self-determination. The former depends on the fact how we view robots' behaviors, and the latter depends on how we should value robots, and thus it is a problem how we should associate with robots within our moral world.

Introduction

Self-determination or the right to self-determination is being talked about in various circles, today. The concepts of self-determination and the right to self-determination have often been used interchangeably, however, it is essential to distinguish the two. Both are distinguished here as follows.

Self-determination is a process involving individual judgments or choices. It depends on circumstances and extends to matters concerning, not only self, but also others. It is affected by other's paternalistic decisions and our paternalistic decisions affect others. At the same time, the *right* to self-determination is a right guaranteed to an individual to make a range of voluntarily decisions without interferences from others or the society. Many bioethical debates continue, for example, on patient's right to self-determination or women's right to abortion. Within a specific range accepted by the society or community, a person has such a right, that is, they are able to decide on matters concerning self.

With this distinction in mind, I want to consider two problems concerning robots. One is whether robots can be self-determining, and the other is whether robots have rights, including the right to self-determination. The topic of robot's self-determination might seem strange, however, the relationship between robots and humans has become closer than ever and I think it is important to look into how humans

interact with robots.

I The robot' s history and the current situation

The word robot was first introduced by Czech playwright Karel Čapek in his play R.U.R. (*Rossum's Universal Robots*) in 1920. It is said that the etymology of the word is from "robota" that means forced labor in Czech. This play was about robots which worked in the robot production company, R.U.R., instead of humans, and rebelled against human society.

As in this play, humans initially produced the industrial-use robot for efficiency. According to the Japan Patent Office (retrieved January 22, 2008, from http://www.jpo.go.jp/shiryou/pdf/gidou-houkoku/robot.pdf), from 1965 to 1980, the first-generation robot was an industrial playback robot that repeated tasks in a predetermined manner. The second-generation robots, from 1980 to 1995, were sensory control robots with a capacity to gather information from the outside world using some sensors and could self-correct their work and behavior to some degree. After 1995, the learning controlled robot appeared. This third-generation robot did not only have a sensory function, but also a learning ability from self experience and had the capacity to reflect it in the next motion. It could also work cooperatively recognizing humans and other robots. Given these new and excellent functions, a whole new range of robots were invented, like extreme environment robots which could work in space, deep sea, disaster sites and mine fields, and service robots which work in medical, welfare, and security environments.

The fourth-generation robot of the 21st century is a type of the artificial intelligence(AI). Now we are entering the age of symbiosis with the robots in the home. For example, the i-Robot Corporation exploited and sold the Roomba, a robotic vacuum cleaner, in 2002 (retrieved January 22, 2008, from http://www.irobot.com/). This robot is currently getting a lot of attention as the first product of the home electrical appliance type. It can perceive bumps and off-limit areas and automatically avoid them or can voluntarily stop moving to avoid coiling of electric codes. It can also find the battery charger and recharge by itself when the cleaning is completed or the battery power decreases.

Another example is Robdog AIBO which was firstly sold in 1999. According to SONY, AIBO has several abilities, for example, the capacity to gather information from the outside world and to independently make decisions based on this information, or, the ability to learn by experience and developing individuality, and the capacity to develop own instinct and

emotion (retrieved January 22, 2008, from http://www.sony.jp/products/Consumer/aibo/). AIBO can play with his own toys by making good use of these abilities.

The AIBO and the Roomba are described as *autonomous* robots according to their developers. What does it mean that they are *autonomous*? Can they decide the direction of movement or destination by themselves?

II Can robots be self-determining?

Here, I will compare humans' self-determination with that of robots. Are both self-determination essentially different? If so, moreover, in what respect do they differ?

Some people may insist that the self-determination is essentially different because only humans have motivation based on emotion and intentionality or because only humans have desires that precede decisions. In contrast, robots have neither desires nor motivations. When the energy decreases, the Roomba does not access the battery charger because it has his desire to re-charge, but because it is only programmed to do so. Though it is said that AIBO has the desire for love and sleep, such desires are not the same as that of man, but just programmed ones.

In response to such opinions, other people may argue that both self-determinations are fundamentally the same. The robots certainly follow a system of input information from the external world, through AI, and output it as some sort of reaction to the world. However, this process is similar to humans. Both humans and robots obtain information through sensors, process it through their data processor (brain in the case of humans), then make an output. In this sense, there is not much difference between humans and robots, and the essential part of the system that decides their actions is the same. In support of this, some researches of brain physiology in recent years suggested that human mental activity can be reduced to electric chemical reactions in our brain. Some people may conclude that our mental phenomenon, our spirit, is ultimately controlled by a program in our brain.

Those people not satisfied with this conclusion might further appeal to "a non-causal free will" of humans. If we do not admit the existence of human non-causal free will, we are accepting that humans are only following a program, similar to robots. On the contrary, if we admit a non-causal free will to humans, there remains a possibility to consider that the human and robots' self-determination are essentially different, that is,

robots are programmed to make certain decisions, although humans can choose other options against the commands of the program (*i.e.* freely). People who hold this view will declare that only humans have free will and human free will is independent of the determinism.

Then, when and how does 'a non-causal free will' exist? Some people may answer that, in human beings, even if decisions were commanded by their power of rationality, we may not follow them. Humans often make bad or wrong decisions, as in the case of yielding to the temptation to drink even if there is still work to be done. It is called human irrationality. Other people may answer that humans can carefully examine a situation and choose the best decision for themselves among various options. In the example above, we can decide to carry on working even if our desire for drinking is very strong. This is human rationality, and it has been traditionally considered that only humans were rational beings. Our self-determination is not ruled by one given program, and has the overall system based on particular phenomena such as irrationality and rationality, and thus humans have been given a special status.

Can robots really not meet these requirements? Some robotics researcher may insist that robots can also overcome these difficulties. They can make robots that have certain irregularities or inclinations in their determination. These robots may hesitate or waver between options just as human do. They can also make (in the future) robots that can make the best choice from many possibilities, like human rationality.

Humans' capacity for making decisions seems to be more superior to that of robots, but this may only be that the human program is currently more or less complex. In other words, the difference between both decisions is only a matter of degree. When we look at an AIBO with a ball, some people think that it is only controlled by a program, but others insist that it is "certainly playing".

This is like two sides of a coin, what is seen depends on the perspective. That is to say, this is a problem of how we see a robot's behavior. It is not a problem that has a "true" or "false", nor should we see it uniformly.

The same thing can also be said about humans. Are humans only controlled by genes or their social imprinting, or are we really free to determine what to do? The answer depends on the person, since how we see a certain person's behavior reflects our attitude to that person, so it follows that, how a robot's behavior is seen reflects our attitude to it. Therefore, we can conclude that a robot decides by itself when it exactly looks like that.

III Can robots have some rights, including the right to self-determination?

The boundary between human beings and machines is becoming increasingly ambiguous. Some robots have higher ability and capacity, for example memory or arithmetic, than humans. A chess computer, *Deepblue*, won a chess game against the world chess champion in 1997. Faced with robots with such an excellent ability, we are compelled to consider that robots can decide by themselves, then, the next problem is that they should be given rights? What rights? The right to live?

We might ask whether robots should be accorded equal value with humans. Many people might still want to say "no" to this problem. Or, the reaction might be that "We should not consider robots like that". Then our discussion will shift from the level of *facts* to the level of *value*, which will lead to the question of how we *should* treat or interact with robots. This discussion is similar to the Person Theory in Bioethics or Speciesism in Environmental Ethics.

The Person Theory is a discussion on what makes a person unique. Michael Tooley, who developed the original Person Theory, took the existence of a long-term desire to continue living as the definition of a person, because he believed that the wrong in killing a person is derived from depriving that person of the long-term and important desire to keep on living. Only existences with such desires can be considered a person and only a person can be considered as a subject with the right to life. Thus, in the Person Theory, a person can be distinguished from the rest of human beings. To have a desire to live, one needs to be competent. According to this, human fetuses or newborn babies, dementia patients and vegetative state patients, and so on, do not qualify to be "a person" in its strictest meaning and cannot be accorded the right to live.

Speciesism is the idea that simply being a human is a good enough reason for humans to have greater moral rights and value than non-human animals. Speciesism is often criticized by people who insist on animal rights. One of them, a modern utilitarian Peter Singer, criticizes Speciesism because he believes that it is morally wrong to discriminate animals simply by virtue of not being humans, just as it is morally wrong to discriminate other people only by the natural differences of race or sex. Instead of differences within species, he emphasizes the presence of the sensory capacity among beings and insists that we should consider the pleasure and pain felt by beings with the sensory capacity. Though he also discriminates living things on the basis of the sensory ability, he believes that this interpretation is more morally correct than others.

As the Person Theory and Speciesism suggest, when we discuss whether or not to attribute rights to certain objects, we are concerned with *values* which they possess. To accord the right to a certain object, it is necessary to admit the qualification and value which it possesses intrinsically. In this regard, how we should identify the various differences among existences becomes a morally important matter.

Going back to the rights of robots, we can identify human uniqueness and this has implications on value hierarchy. By such a measure, we can try to interpret particular features that are used to confer human status, to machines. In such a process, we often enumerate human rationality or practical reason as a positive value. At the same time, because it becomes possible to make robots similar to humans now, to emphasize the essential difference from machines, we often enumerate irrationality, fatigue, aging or death in humans as our negative value.

Finally, from the viewpoint of moral world, if we think we should treat robots in the same way as humans, we have to accept that there is some value to robots as 'moral agents' and they possess some rights. On the contrary, if we do not want robots to actively participate in our morality, then they are not admitted the status as a moral agents, and will not have their rights. So even if we select the latter, this does not justify that we need not consider robots at all. We still might have to care about robots morally. In that case, robots could become objects of our moral consideration even if they are not moral subjects.

If robots are considered as such, we notice that our consideration for robots approximates that of human embryos and fetuses, animals, nature and ecosystems. We are unlikely to clearly declare that these existences have some rights like humans as moral agents, however, those beings are alive with us in the same world. Some robots, especially AIBO and Roomba, are also certainly have established a close relationship with us. If it is certain that we are moral agents, at least, we could somehow care for beings which might not possess rights.

Conclusion

How should we associate with robots? Our answers to this question will determine what kind of robots we should make. Various answers can be proposed. One is that we maintain that robots are just a mere tool to help us and make our lives more convenient. Another is that we should consider robots as objects for further scientific exploration, or as partners living with us. It influences the problem of robot' s rights according to our answer to the question.